



# GLOBAL SHIELD.

## Press Release

### An AI Safety Institute is a strong first step, concrete regulation is still needed to ensure its success

*Devon Whittle, Global Shield's Australia Director: "We recently showed MPs how easy it is to weaponise artificial intelligence, and welcome the government taking this strong step towards addressing AI risk by establishing an Australian AI Safety Institute (AISI). This needs to be followed by concrete regulation to support the AISI's work and protect Australians."*

Canberra, 25 November 2025 – Global Shield Australia welcomes the announcement of an Artificial Intelligence Safety Institute (AISI), as an important first step to getting Australia's regulatory settings right on AI.

The AISI will provide the Australian government with the critical technical expertise needed to identify and assess AI threats and inform the government's response. It will also help Australia to engage with and lead regional and global efforts on responsible AI development.

To be effective, the AISI needs to be supported by transparency and reporting obligations on AI companies, in line with jurisdictions such as the European Union and California. It should also be followed by concrete obligations on AI companies to reduce the risk from their products.

Recently, at a private briefing in Parliament House, Global Shield showed MPs how easily off-the-shelf AI models can be used to threaten Australia's democracy and national security. From cloning politicians' voices to generating instructions for making a virulent pathogen, MPs saw the dark side of this world-changing technology.

"The demonstration revealed the gap between Australia's lack of AI regulations and the harm that can be caused by the malicious use of today's AI models. The government's announcement of an AISI is a strong first step towards bridging that gap. We look forward to it being followed by concrete reforms to ensure that our regulators can keep pace," said Devon Whittle, Global Shield's Australia Director.

### ***Voluntary, fragmented safeguards mean Australians are less safe***

Current safeguards on AI models are voluntary, fragmented, and reactive. This means that some companies apply some safety measures, some of the time. It also means that without action, Australians will be less protected than those overseas.

For example, a popular online platform allows the voices of Australian politicians to be cloned in seconds. But that same platform refuses to clone certain American political figures. The unlevel playing field means that some Americans are protected, while Australians are exposed.

### ***What Australia can do to address AI risk***

The EU's AI Act imposes mandatory safeguards on high-risk systems. California has transparency and incident reporting requirements, and in our region, Malaysia and Vietnam are pushing ahead with concrete regulation.

Similar measures in Australia are needed to build upon the foundation that the AISI will provide. This means taking steps to bring AI regulation in line with every other high-risk industry.

- Firstly, we need to make large AI companies responsible for the safety of their products, starting with measures to prevent truly catastrophic outcomes.
- Secondly, we need mandatory incident reporting, just as we have for sectors such as pharmaceuticals and aviation. When advanced AI malfunctions or causes serious harm, the government should know. The AISI would be the right body to receive these reports.
- Finally, we must back an Australian AI assurance industry, to give consumers and businesses confidence that products are actually safe and secure. We have mandatory cyber security labelling of smart speakers, but little for far more powerful AI tools.

These first steps would ensure Australians have at least comparable protections to their counterparts overseas and position us to lead on responsible AI in the region. They would also unlock the potential economic benefits of AI by building trust in this technology amongst Australians.

ENDS

#### **About Global Shield**

Global Shield is an independent non-profit advocacy organisation dedicated to reducing global catastrophic risk. Our mission is to help governments and leaders enact and implement effective policies to reduce the risk of global catastrophe from all hazards.

Contact us: [australia@globalshieldpolicy.org](mailto:australia@globalshieldpolicy.org) / [marvin.meintjies@globalshieldpolicy.org](mailto:marvin.meintjies@globalshieldpolicy.org)